

DOCUMENT RESUME

ED 241 554

TM 840 059

AUTHOR Lunneborg, Clifford E.
TITLE Efron's Bootstrap with Some Applications in Psychology.
PUB DATE Aug 83
NOTE 8p.; Paper presented at the Annual Meeting of the American Psychological Association (91st, Anaheim, CA, August 26-30, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Error of Measurement; *Estimation (Mathematics); *Hypothesis Testing; Regression (Statistics); Research Methodology; Sampling; *Statistical Analysis; *Statistical Distributions
IDENTIFIERS Efron (Bradley); *Efrons Bootstrap; General Linear Model; *Inference (Statistical); Repeated Measures Design

ABSTRACT

The wide availability of large amounts of inexpensive computing power has encouraged statisticians to explore many approaches to a basis for inference. This paper presents one such "computer-intensive" approach: the bootstrap of Bradley Efron. This methodology fits between the cases where it is assumed that the form of the distribution is known and cases where the sample provides no information about the form of the population distribution. A series of samples are drawn, randomly and with replacement. From each sample, the value of a parameter is estimated. The bootstrap conjecture is that the sampling distribution of the statistic being studied and sampling distribution found from this iterative process are essentially identical for a wide variety of statistics. Some areas of application for this approach include regression and fixed effects analysis of variance, the median general linear model, the repeated measures design, and errors of measurement. (BW)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED241554

* This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

* Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

Efron's Bootstrap with Some Applications in Psychology

Clifford E. Lunneborg
University of Washington

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

C. Lunneborg

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Introduction

The statistics psychologists depend upon most heavily for drawing inferences or making generalizations from a sample of observations are parametric and almost exclusively normal theoretic. We test hypotheses in the analysis of variance, evaluate the significance of a correlation coefficient, or establish confidence bounds for the magnitude of some effect of interest by virtue of having assumed our observations to have been drawn from populations which are Gaussian in form. By making the assumption we are able, relatively easily, to calculate what is necessary for decision or estimation.

Where the normal theory assumptions are clearly inappropriate (and we lack assurance that our normal calculations are 'robust') we fall back on one or another nonparametric approach. To date, the popular nonparametric procedures rely upon rank statistics. In effect, we can replace our observations with their ranks and continue the analysis. With some consequent loss of inferential power, then, we may choose to go from assuming that we know a great deal about the form of a distribution (i.e., that is normal) to assuming that we know essentially nothing about it (i.e., all we can do is to order observations). Again, the alternative is appealing because we can calculate, straightforwardly if not easily, the necessary rank order statistics.

The wide availability of large amounts of inexpensive computing power has encouraged statisticians to explore other approaches to a basis for inference. It is possible, they are beginning to argue, that computers can both advance statistical theory and provide considerably better methods of solution where no good theory exists (Efron, 1979b and Note 2). It will be the purpose of this submission to present one such 'computer-intensive' approach, suggesting areas of application in psychology. The approach is the bootstrap of Bradley Efron (1979a, 1982 and Diaconis and Efron, 1983). His proposed methodology is important because it promises to be widely applicable and it fits neatly between the alternatives presently available -while we need not assume that we know the form of the distribution from which we are sampling, we also need not assume that our sample, once collected, gives us no information about the parental form.

The Bootstrap Principle

The bootstrap is perhaps most easily understood by first developing a little general statistical inference. Let X be a population (of potential observations) characterized by some parameter θ . Next, let x_1 be a (random

and independent) sample of n observations drawn from X . From x_1 we calculate S_1 , an estimate of θ .

How close is S_1 to θ ? Classical statistical inference teaches that S_1 - or $(S_1 - \theta)$ - is an observation from a sampling distribution. That is, we could in principle generate from X an unlimited series of independent samples, each of size n , $x_1 = x_1, x_2, x_3, \dots$, and from each of these calculate an estimate of θ , $S_i = S_1, S_2, S_3, \dots$. The collection of values S_i - or $(S_i - \theta)$ - constitutes a sampling distribution.

Let F designate the distribution of $(S_i - \theta)$ approached as the number of estimates increases without limit. We use information about F - its mean, variance, and shape - to test hypotheses about θ or to assess its estimator. [The advantage of normal theory statistics, of course, is that we don't have to repeat our study over and over to create F . If X is normally distributed then we know what F has to look like.]

We continue with $X, \theta, x (= x_1)$, and $S (= S_1)$ defined as above. From the collection of observations comprising x we now sample n times, randomly and with replacement. The result constitutes a first bootstrap sample, x_1 , from which we calculate a first bootstrap estimate, S_1 . [S_1 is calculated from x_1 in exactly the same fashion as S is from x .] Repeating the same process that gave us x_1 , we can create a sequence of bootstrap samples, $x_1, x_2, x_3, \dots$, and bootstrap estimates, $S_1, S_2, S_3, \dots$. [Each x_i consists of n draws from x , made randomly and with replacement.] We let F^* designate the (bootstrap) distribution of $(S_i^* - S)$ for some 'large' number of bootstrap estimates.

F is the sampling distribution of the statistic S and reflects repeated sampling from X . F^* is the sampling distribution of S^* and reflects repeated sampling from x . The bootstrap conjecture (Efron, 1979a, 1982) is that F and F^* are essentially identical for a wide variety of interesting statistics. The importance of this is that the characteristics of F^* (and, hence, of F) may be established (a) with no assumptions about the form of X , (b) with no repeated sampling (experiments) from X , but (c) with (perhaps considerable) resampling from x and re-computing of statistics of interest.

Some Bootstrap Results

It is early days for judging the potential of the bootstrap principle, but the first results are encouraging. Efron (Note 2 and 1982) has demonstrated, theoretically and via simulation, that the bootstrap can provide a more accurate picture of the variability of an estimate than do other 'resampling' procedures (e.g., Tukey's jackknife). Freedman and Bickel (Bickel and Freedman, 1981 and Note 1 and Freedman, 1981) have established the validity of the bootstrap conjecture for two important classes of problems - regression and correlation.

The early studies suggest that certain statistics, or forms of these statistics, are more successfully 'bootstrapped' than are others. Symmetry of distribution seems important: the mean (or a linear function of means) works well but $\log S^2$ performs better than S^2 and $\tanh^{-1}(r)$ (or, Fisher's z transform) better than r itself. Finally, the first results also give some idea of how many bootstrap samples may be necessary to tell us what we want to know about F^* (and, hence, F). As few as 100 bootstrap replications can provide a stable estimate of the standard deviation of the sampling distribution of a statistic but a tenfold increase in replications may be needed for accurate 'nonparametric confidence intervals' (Efron, 1980).

The decision to use a bootstrap approach involves a trade-off - between the cost of computation and the accuracy of results. Where computational resources are available and of minimal marginal cost - e.g., a microcomputer which might be allowed to run unattended overnight - and a highly accurate estimate of, say, an effect size is needed, the bootstrap could play an

important role. [With the proliferation over the next 2-5 years of microcomputers which compute considerably faster than do the present generation -the MC68000 microprocessor, at the heart of several new 16 bit machines, executes 2 million instructions per second- the choice of a computational approach will be even more appealing.]

Some Areas of Application for the Bootstrap

For certain of these areas, notably the first, bootstrap theory is well enough advanced that the application is quite straightforward. For others, notably the last, we can only speculate on the appropriate bootstrap approach.

Regression and Fixed Effects ANOVA. The bootstrap obviates the assumption of normality in these designs. It is now known (Freedman, 1981) that the appropriate bootstrap procedure is to resample 'errors' rather than cases. If a separate bootstrap distribution is developed for each hypothesis tested in the ANOVA (i.e., for each contrast among population means), the lack of independence associated with hypothesis testing in the unbalanced design may be overcome.

The Median General Linear Model. The above bootstrap approach to regression, the most thoroughly studied one, assumes a homoscedastic error model. Where lack of homogeneity is appropriate, more radical bootstrap approaches are possible. One of these would substitute linear contrasts among population medians for contrasts among means. Not only are medians effectively bootstrapped (Efron, 1979a) but the bootstrapping yields a theoretically determinate structure -Monte Carlo sampling is not required, although the analysis is still 'computer intensive'.

The Repeated Measures Design. The analysis of repeated measures via classical analysis of variance requires not only normality and homogeneity of variance but that the variance-covariance matrix take a special form -compound symmetric or less restrictively, 'Type H'. These assumptions may be difficult to meet and our techniques for detecting whether or not they are met are decidedly fallible. Bootstrapping could provide assessments of 'within S effects' with which we would be more comfortable. [See Appendix for one possible bootstrap approach to a repeated measures problem.]

Errors of Measurement. Classical psychometric theory interprets a standard error of measurement, common for all examinees, based on normal theoretic considerations. Modern (latent trait) test theory allows for different error bounds for different score patterns but these are based on item calibrations that are strongly parametric. Bootstrapping individual (and item) performance may provide more realistic assessments.

Reference Notes

1. Bickel, P. and Freedman, D. More on bootstrapping regression models. Technical Report, Department of Statistics, University of California, Berkeley. 1981.
2. Efron, B. Computer intensive methods in statistics. Technical Report No. 150, Department of Statistics, Stanford University, Stanford, CA. 1980.

References

- Bickel, P. and Freedman, D. Some asymptotic theory for the bootstrap. Annals of Statistics, 9, 1196-1217, 1981.
- Diaconis, P. and Efron, B. In Scientific American, May 1983.
- Efron, B. Bootstrap methods: another look at the jackknife. Annals of Statistics, 7, 1-26, 1979a.

- Efron, B. Computers and the theory of statistics: thinking the unthinkable. SIAM Review, 21, 460-480, 1979b.
- Efron, B. Censored data and the bootstrap. J. Amer. Statist. Assoc., 76, 312-319, 1980.
- Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA. 1982.
- Freedman, D.A. Bootstrapping regression models. Annals of Statistics, 9, 1218-1228, 1981.

Appendix

(On Bootstrapping the Sobering Sailors)

We have data of the following kind: After variable periods of initial delay (since last drink) Ss are administered a battery of cognitive tasks. The battery is readministered (at fixed intervals) a number of times.

Although the battery consists of a number of tasks, we will concentrate on the analysis of performance on a single task.

To represent the problem we will assume k different delays to initial testing, defining k groups of Ss, and m repetitions of the test battery. The total number of Ss is n , resulting in $N = n \times m$ observations of task performance. For certain aspects of the model development, notably the construction of a sampling model (in the form of an empirical error distribution) we will regard the m repetitions as constituting a multivariate observation on each of the n Ss. For other aspects, notably the representation of interesting determinant or structural models, we will regard the set of $N = n \times m$ test scores, each a subject-occasion combination, as univariate observations-- separate, though not statistically independent.

The goal is to represent task performance (y) as a function of time since last drink (t), number of repetitions of the task (r), and-- perhaps--an interaction between the two.

$$y = f(t) + g(r) + h(t, r)$$

The functions f , g , or h may be polynomial or they may be non-linear functions of r or t (i.e., we may need to transform t , r or their product before writing the model).

There may be competing models of the type above. If we can arrange these hierarchically we can argue for the sufficiency of one over the other(s) on the basis of the sampling variability (here, bootstrap) of the estimators of the parameters in the models. If competing models are non hierarchical the argument is that one (or all but one) imply parameters with 'bad' sampling distributions.

The problem arises in that (1) the models of interest are not standard repeated measures ANOVA representations (2) nor are the observations y independent in the sense required by the standard linear regression or GLM model.

Pull on the bootstrap.

Part 1. Developing the Empirical Error Distribution for Bootstrap Sampling

a. We let X be the $n \times k$ design matrix, coding group membership of the n S s, and take Y to be the $n \times m$ criterion matrix, the i th row consisting of the m repetitions of task performance for the i th S . The usual MGLM structural model is given by

$$E(Y) = X \beta$$

or

$$Y = X \beta + \varepsilon$$

where β is the $k \times m$ matrix of (unknown) population regression coefficients and ε is a $n \times m$ matrix of (unobserved) 'errors'. Each row of ε is a multivariate observation from a population with expectation 0 (the m element null vector) and variance σ , a $m \times m$ product moment matrix.

b. We use the usual (least squares or normal maximum likelihood) estimator of β :

$$\hat{\beta} = (X'X)^{-1} X'Y \quad [1]$$

to give us a matrix of 'observed' (or, estimated) errors

$$E = Y - X \hat{\beta}$$

or

$$Y = X \hat{\beta} + E \quad [2]$$

(Note: $\hat{\beta}$ is not itself of interest, nor is $X \hat{\beta}$. $\hat{E}(Y) = X \hat{\beta}$, though, is. It will be, simply, a matrix of sample averages, an average being computed for each of the k groups on each of the m occasions.)

c. We will use E as the empirical 'stand in' for the unknown distribution of error vectors for our bootstrapping.

Part 2. Getting the Structural Model Parameter Estimates

a. Form the Y matrix above into a N ($= n \times m$) element y vector by concatenating the rows together. Then, construct the $N \times p$ matrix W so the columns of W contain the appropriate functions of t , r , and $t \cdot r$. (Note: If competing models are not hierarchical there may be more than one W matrix.)

b. Compute

$$G' = (W'W)^{-1}W'$$

[3]

and

$$a = G'y$$

[4]

The p element vector a provides point estimators of our theoretically interesting parameters. We will bootstrap to find out something of their sampling variability.

(Note: If there is more than one W matrix to cope with you can replace G with

$$\gamma = ({}_1G, {}_2G, \dots)$$

both in [4] above and in the section following.)

Part 3 The Bootstrap Process

a. Each bootstrap trial will require the matrices E and $E(Y) = X\beta$ from [2] and G from [3].

b. Here is the bootstrap. Add to each row of $E(Y)$ in turn a randomly chosen row of Z . Concatenate the resulting rows together to produce the vector

$${}_jY^*$$

where the subscript tells us that this vector belongs to the j th bootstrap trial.

c. Compute

$${}_j a^* = G' {}_j y^*$$

[5]

This vector provides the set of j th elements for the bootstrap sampling distribution of the p a parameters.

d. Arrange to save and describe the p bootstrap sampling distributions as j ranges over $1, 2, 3, \dots, J$ where choice of J is from the set $[100 \leq J \leq 1000]$.

Computational Note: To save memory in our microcomputer implementation (in Apple Pascal and CP/M Fortran-80) computations have been made dependent upon ${}^T W$ and $H' = (W'W)^{-1} {}^T W'$ rather than upon W and G' . W is reduced to ${}^T W$ by noting that the typical row of W appears n_i times.